Biophysical Chemistry, 32 (1988) 305–317
Elsevier

BPC 01335

A theoretical model of DNA curvature

P. De Santis [a], A. Palleschi [b], M. Savino [a] and A. Scipioni [a]

[a] *Dipartimento di Chimica, Dipartimento di Genetica e Biologia Molecolare, Università di Roma "La Sapienza", P. le A. Moro 5, 00185 Roma, Italy and* [b] *Istituto di Chimica, Università della Basilicata, Via N. Sauro 85, 85100 Potenza, Italy*

Received 3 August 1988
Revised manuscript received 27 October 1988
Accepted 3 November 1988

DNA curvature; Conformational energy calculation

Distortions from the uniform idealized B-DNA structure are investigated in terms of differential interactions between adjacent nucleotide pairs on the basis of conformational energy calculations. A theoretical model of DNA curvature is proposed based on the evaluation of the curvature vector defined in the complex plane and the corresponding variance. The model appears to contain the basic physical features for translating the deterministic fluctuations of DNA sequences in superstructure elements. It allows the quantitative reproduction of all the available gel electrophoresis experiments on both periodical polynucleotides and tracts of DNAs as well as the theoretical prediction of the sequence dependent DNA writhing in good agreement with the experimental data. The general pattern of agreement between the theoretical and experimental data and the biological significance of the results obtained allow an extensive application of the model for the screening of DNA regions which are possible candidates for protein recognition.

1. Introduction

It is now becoming a general concept as a result of intensive investigations of the last few years, that DNA is characterized by intrinsic sequence dependent superstructures which are suggested to be relevant for a number of gene control mechanisms and recognition processes.

The first hypothesis about the possible factors responsible for the ability of DNA to translate, at the stereochemical level, the slight differential interactions of the base pairs along the sequence is due to Trifonov [1,2], who suggested that wedge formation in base pair stacking (mainly localized at AA (TT) dimers in the tilt direction) could be the cause of the bend of the DNA double helical axis where it is distributed in phase with the period of the DNA structure. The wedge model was proposed by the author to give a significance to the observed base pair fluctuations in the eucaryotic DNA in relation to the question of the preferential positioning of nucleosomes in DNA.

Some regularities in the distortions from the ideal uniform standard B-DNA structure were observed in double helical fragments of polynucleotides investigated by X-ray crystal analysis in spite of the packing perturbations and terminal effects [3,4].

In recent years anomalies in the electrophoretic pattern were observed in natural DNAs [5–9] and interpreted as a manifestation of the presence of stationary curved double helix regions. Both X-ray and electrophoretic experiments, supplemented by gel filtration and rotational relaxation measurements, convincingly proved the ability of DNA to distort its helical axis under sequence dependent stereochemical restraints.

The term 'curvature' is, at present, assumed to indicate such an intrinsic property of DNA, whereas the term 'bend' is used to define the ability of DNA to anisotropically deform under the external constraints of protein interactions.

Correspondence address: P. De Santis, Dipartimento di Chimica, Università di Roma 'La Sapienza', P. le A. Moro 5, 00185 Roma, Italy.

Stationary and dynamic distortions of the B-DNA structure were, very recently, detected in synthetic double helical oligonucleotides by the interpretation of NMR data [10,11].

Finally, the existence of significant sequence dependent stacking perturbations and some evaluations of deviations from the ideal B-DNA were obtained on the basis of conformational energy calculations by different authors [12–15].

Although generally accepted, the curvature of DNA is a current subject of debate and the object of intense investigations to clarify the stereochemical factors which are determinant. Systematic electrophoretic experiments by Koo et al. [16], Hagerman [17] and Diekman [7,18] showed that periodical double helical polynucleotides obtained by ligating mainly synthetic decanucleotides, but also some oligonucleotides with different periods (9, 11, 12, 15, 20, 21), cover a large interval of retardation values (corresponding to the ratio between the apparent and the true molecular weight) between 1 and 2.7, in some logical connection with the sequences. This gave rise to some qualitative explanations about the origin of the curvature which can be recognized as belonging to two models: one localizes the curvature at a hypothetical structural transition between the B-DNA structure [16,19] and the so-called H-DNA (heteronymous) structure which an interpretation of the fiber X-ray diffraction pattern of poly(dA)-poly(dT) assigned to -AA- repeated sequence (more than 3) [16]; the other one derives the curvature of DNA from a constructive summation of local distortions of nearest neighbor base pairs [1,2].

We adopted the latter hypothesis, as a first approximation, and proposed an original model based on the theoretical evaluation of the local deviations from the idealized B-DNA structure of the 16 different stacked base pairs. The model was able to reproduce very satisfactorily the experimental electrophoretic data available at that time on periodical polynucleotides and to predict the localization of the curvature in natural DNAs in agreement with the experimental data [14,15].

The present paper is a refinement of the model proposed earlier; it introduces new aspects of the problem and useful representations which allow a surprisingly good quantitative theoretical reproduction of all the available gel electrophoresis data concerning the periodical polynucleotides as well as the fine features of the permutation gel electrophoresis experiments on natural DNAs; the model also allows an easy localization of curved DNA regions and the theoretical prediction of the main features of DNA superstructures from the sequence as proved by the strikingly good direct comparison with their electron microscopy visualization [21]; finally, it appears suitable for an extensive screening of the DNA regions which are possible candidates for protein recognition.

2. A theoretical model for predicting DNA superstructures from the sequence

The model assumes that the local deformations of the dinucleotide double helix fragments are in first approximation a useful basis, good enough for predicting the superstructural features of DNAs and their experimental manifestations using appropriate calculation methods.

The local deviations of the different dinucleotide fragments were obtained by the methods of the theoretical conformational analysis we first introduced about 25 years ago [22,23], using energy minimization procedures in the multidimensional conformational space and our potential energy set of functions which allowed reproduction of the fine features of the standard DNA structures [24–26] as well as prediction of the conformations of polypeptide chains and complex natural peptides in good agreement with experimental X-ray structures [27,28].

The integration of the local structural deviations from the ideal canonical B-DNA structure was obtained as illustrated in the appendix in terms of the curvature vector $C(n, v)$ which represents in the complex plane (in modulus and phase) the directional change of the double helix axis between the sequence number n and $n+v$. It is given per turn of B-DNA by:

$$C(n, v) = v^0/v \sum_{j=n}^{n+v} d_j \exp[2\pi i(j-1)/v^0]$$

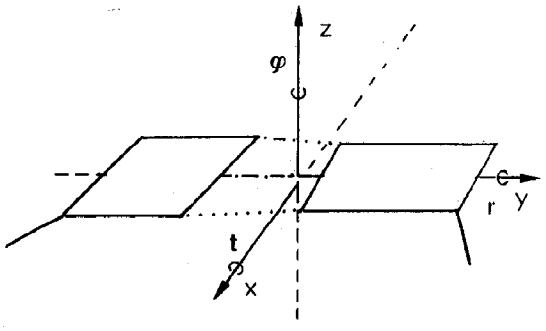


Fig. 1. Base pair orientational parameters.

where v^0 is the average periodicity of DNA = 10.4 and $d_j = (r_j - it_j)$ (r_j and t_j the roll and tilt angles respectively, as defined in fig. 1) represents the orientational deviations of the jth base pair average plane from the canonical B-DNA. It should be noted that only the average twist of the tract of DNA considered sensitively affects the DNA superstructure and only when v is larger than a few turns of DNA.

The appendix illustrates the concept of the curvature vector, which is well known in the elasticity theory of rod flexure, as obtained from the transformation matrices at the first order of the pertinent Taylor series; the same final formula was, however, obtained using the theory of signals and group theory (to be published).

The curvature vector is represented by the Fourier transform component of frequency equal to the B-DNA periodicity, v^0, of the local structural fluctuations along the sequence d_j. It is about zero unless the structural fluctuations contain a harmonic component of period v^0; further it is invariant if all d_j are changed for a constant value or if the distribution of the local deformations is changed but saving the component of v^0 periodicity of the corresponding Fourier series. Finally, it is interesting that whilst the formula of the curvature vector as derived from the transformation matrices holds at the limit of the small angles, namely, for slight local deviations from the canonical B-DNA, neverthless, because of its symmetry properties, it allows build up of superstructures of DNAs very close to those obtained with the full matrix transformations, also for local deviations that largely overcome the limits of the linear terms of the Taylor series (which we have assumed in order to derive the given formula for the curvature vector).

3. Theoretical evaluation of local distortions on the basis of conformational energy calculations

The problem of the DNA curvature was then reduced to the evaluation of the deviation vectors, d_j, namely of the roll and tilt angles, for the different base pair dinucleotide fragments of DNA. These can be obtained from the X-ray structures of relevant double helical oligonucleotides where some regularities were observed and coded in the Calladine-Dickerson rule [29].

However, some doubts still remain about the significance of the deformations observed near the terminals as well as about the question of the transferability of the crystal structures in the biological conditions in solution, where the distribution of water molecules and counterions is plausibly different as well as the whole molecular dynamics.

Also, the NMR results on double helical oligonucleotides are a useful source of data about the sequence dependent distortions. However, apart from some pending questions about the NOE data interpretation, the molecular dynamics effects near the terminals probably influence the local distortions, which appear different from the X-ray data.

We tried to obtain theoretically the dinucleotide distortions by localizing the minimum of the conformational energy, using a constrained minimization procedure which takes into account the stereochemical conditions of the correct formation of base pairing with a fixed propeller twist of 18^0 and 12^0 for A-T and G-C pairs according to the experimental average values observed in relevant X-ray structures [30]. We adopted our set of potential functions [25,26] but provided a larger van der Waals radius of 0.25 Å to the methyl group to mimic hydrophobic effects [14,15].

The results of such calculations are reported in the following matrix form for the roll and tilt angles:

r rad	A	T	G	C
T	−0.16	0.09	−0.12	−0.04
A	0.09	0.12	0.02	0.04
C	−0.12	0.02	−0.10	−0.01
G	−0.04	0.04	−0.01	0.07

t rad	A	T	G	C
T	0.00	0.00	0.01	−0.03
A	0.00	0.00	0.03	0.03
C	−0.01	−0.03	0.00	−0.02
G	0.03	−0.03	0.02	0.00

Note the dominant role of the roll distorsion and the symmetry and the antisymmetry of the roll and the tilt matrices, respectively, which ensure that the same curvature is obtained using the sequences of both the complementary strands. Further, the theoretical matrices appear to account for the Calladine-Dickerson rule [29], proposed on the basis of crystallographic data on relevant model compounds as proved by the alternance of the roll values of Pyr-Pur and Pur-Pyr dinucleotides; and finally, they satisfactorily account for the trends of the distortions along the sequence in the crystal structures of dodecamers as recently reviewed by Dickerson [30] (see fig. 2).

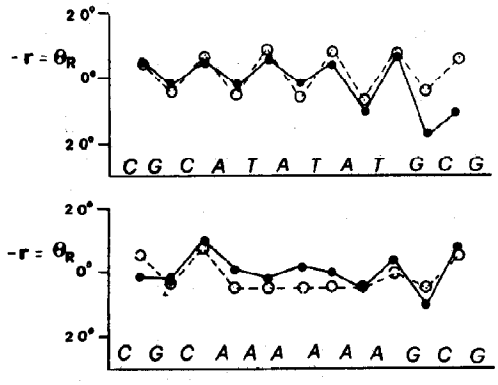


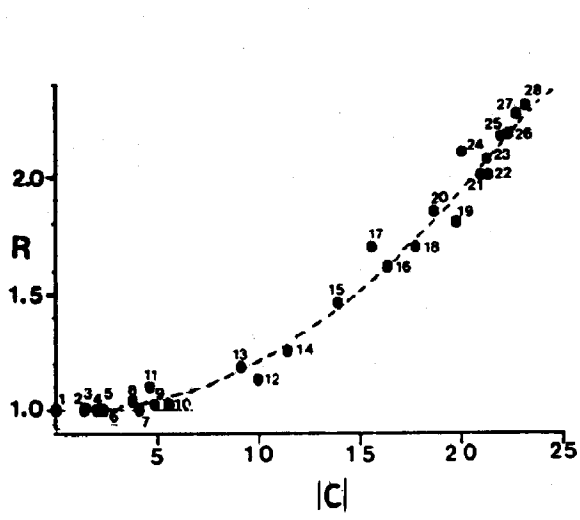
Fig. 2. Comparison of the roll angles obtained by conformational analysis, r (○) with those obtained by X-ray diffraction analysis, θ_R (●) (the sign of r is inverted because of a different convention adopted) [30].

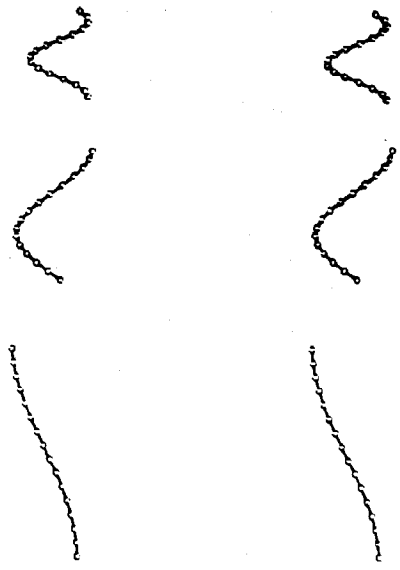
4. Correlation between curvature and electrophoretic retardation in periodical polynucleotides

The curvature was calculated in the case of the polynucleotides with period 10, 11, 15 and 20 synthesized by Koo et al. [16], Hagerman [17] and Diekmann [7,18] as well in our laboratory (to be published); their retardation factors were reported versus the modulus of the curvature as illustrated in fig. 3 where the representative points of polynucleotides of 150 bp are shown. The diagram appears very similar to those previously published by us [14,15] but the abscissa, now, represents a good approximation of the curvature per turn in degrees. The trend is similar to that of the electrophoretic retardation versus the molecular weight reported by the cited authors, suggesting a similar functional dependence of the retardation by either molecular weight or curvature. A striking result is that the model correctly predicts the dramatic difference of the electrophoretic retardation observed for the two pairs of the Hagerman polynucleotides CA_4-GT_4 and GA_4-CT_4, which are related by a simple inversion of the sequence. In the same figure the stereoscopic projections of the writhing of polydecanucleotides with high, medium and low curvature are shown; note the left-handed screw sense as a consequence of the negative value of δ ($= -0.24$ rad), the phase of the repeating oligonucleotide unit along the B-DNA structure.

However, when polynucleotides with different periodicity were considered, the corresponding representative points resulted in general disagreement with the theoretical curve of fig. 3, indicating the importance of the phase. Therefore, we considered this factor in a first approximation by correcting the values of curvature for the cosine of the pitch angle of the superhelical axis, evaluated as $|C|/(|e|^2 + \sin\delta)^{1/2}$.

Such an effective curvature, corresponding to $C \cdot |C|/(|C|^2 + \sin^2\delta)^{1/2}$, tends to C when $\delta \to 0$ and rapidly decreases when δ becomes greater than $|C|$. As a result, polynucleotides with periodicity different from 10 or 11 ($\delta = -0.24$ and 0.36 rad respectively) are characterized by very low effective curvature and practically behave as normal DNA in gel electrophoresis. In spite of the approximate nature of such a correction, the cor-





1) A5N10	CCGGCAAAAACGGGC	15) CG2A2	CGGAATTCCG
2) C2T3	CCTTTAAAGG	16) A4N6	GGCCAAAACG
3) TC4	TCCCCGGGGA	17) A9N1	CAAAAAAAAA
4) AG4	AGGGGCCCCT	18) A5-8	CCAAAAACGGGCAAAAAAAA
5) AC4	ACCCCGGGGT	19) FGG	CCGAAAAAGG
6) TG4	TGGGGCCCCA	20) A8-5	CCAAAAAAAACGGGCAAAAA
7) C3G3	GGGTCGACCC	21) A5N5	GGCAAAAACG
8) G5N5	TCGTGGGGGC	22) G2A3	GGAAATTTCC
9) IAC	GGCAACAACG	23) FCT	GGCAAAAATG
10) CT4	CTTTTAAAAG	24) A8N2	CCAAAAAAAA
11) GT4	GTTTTAAAAC	25) GA4	GAAAATTTTC
12) IAG	GGCAAGAACG	26) A6T6	AAAACGGGTTTTTTGGGCAA
13) IAT	GGCAATAACG	27) CA4	CAAAATTTTG
14) A3N7	GGCCAAACCG	28) A6N4	GGCAAAAAAC

Fig. 3. Correlation diagram between the electrophoretic retardation of the sequential polynucleotides synthesized by Hagerman [17], Diekmann [7,18] and Crothers et al. [16] and the modulus of the curvature vector $|C|$. The superhelical structures corresponding to $|C| = 25^0$, 15^0 and 5^0 are reported as stereoprojections.

relation between experimental retardation and the effective curvature appeared very satisfactory [31].

5. Curvature function and DNA superstructures

The model was extended to describe the curvature of natural DNAs along the sequence. It is conveniently represented by a pair of diagrams where both the modulus and the relative phase (calculated from the first nucleotide residue) of the curvature vector $C(n, v)$ are reported versus the sequence number n. We assumed $v = 31$, i.e., about 3 turns of DNA ($v^0 = 10.4$) and assigned the value of the curvature to the sequence number $(n + 15)$, namely to the center of the sequence tract considered; such diagrams provide a powerful tool for localizing the DNA curvature along the sequence and for obtaining the DNA writhing in the space. In fact, the modulus of curvature can be assumed as the virtual bond angle between the vectors representing the local helical axis of two

successive turns of DNA and the increment of the associated phase angles as the corresponding torsional angle around the virtual bond; these are the internal parameters to define the writhing of DNA axis.

We have evaluated such a curvature function in several DNA fragments characterized by unusual features of the electrophoretic properties. Fig. 4 shows the curvature function of a 659 bp fragment of the plasmid pPK201 containing 221 (275–495) bp of *Crithidia fasciculata* minicircle, which has been shown to have a very anomalous electrophoretic mobility [32]. The diagram clearly shows a DNA tract characterized by both a high curvature and a constant phase from which a complete superhelical turn can be predicted. In fact, the top of the figure shows the stereoprojection of the DNA writhing as obtained from the modulus of curvature and the phase increment per turn of B-DNA.

It is exciting that the theoretically predicted DNA loop was actually localized in that position by electron microscopy [21], which revealed 200–300 bp loops in relaxed as well as nicked

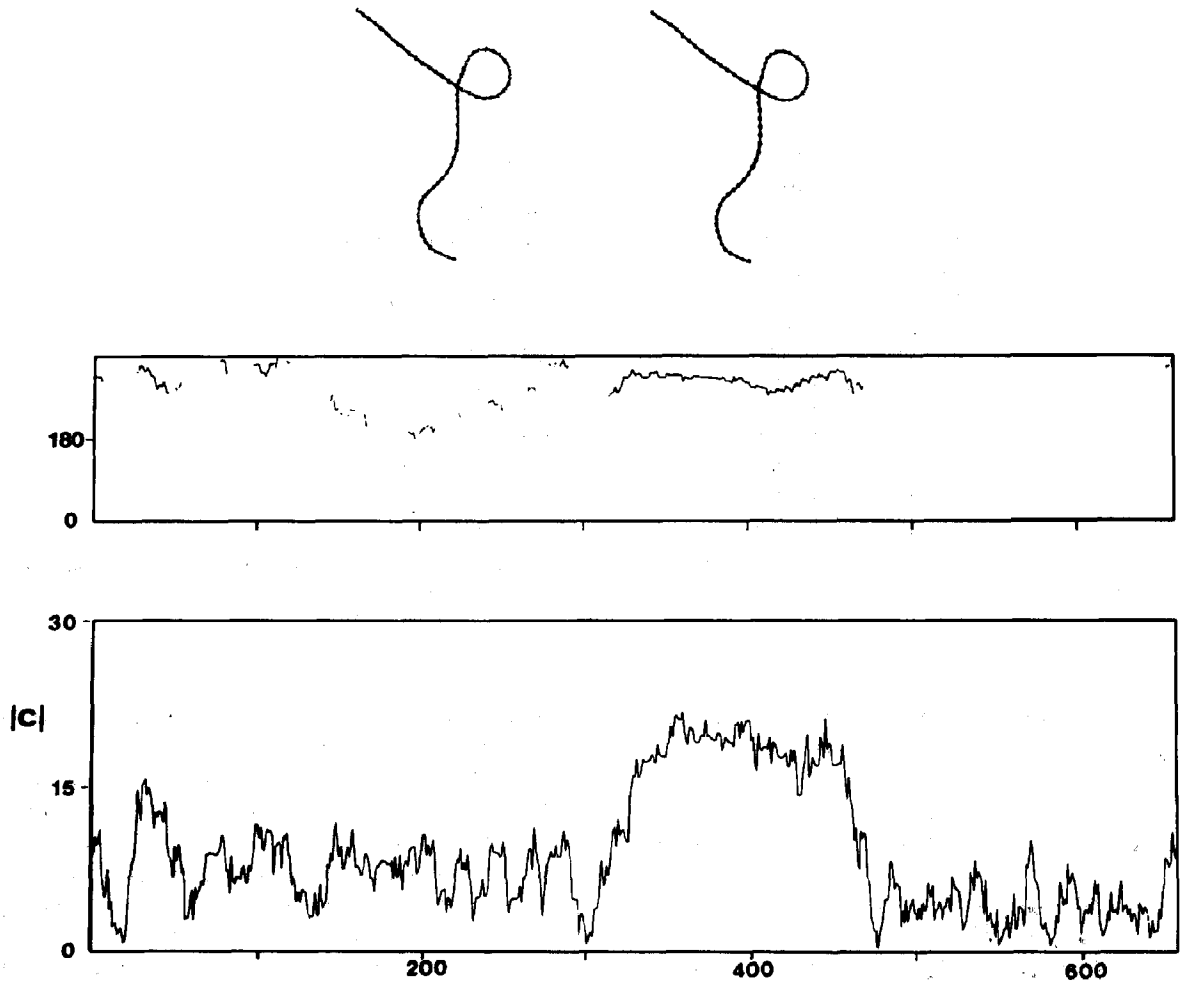


Fig. 4. Curvature profile and relative phase of the 569 bp DNA cloned fragment of *Crithidia fasciculata*; the writhing of the double helix axis shown of a stereo projection, forms a loop as visualized by electron microscopy [21].

minicircle (2.5 kb) and showed circular forms in up to 70% of the observed molecules in the 219 bp DNA fragment containing the sequence 275–495.

Thus, in the case of 490 bp kinetoplast DNA from *Leishmania tarentolae*, the curvature diagrams clearly show the presence of a prominent maximum which is characterized, also, by the practical invariance of the phase. This indicates the existence of a main flat curvature with center at $n = 136$ and a global angle of about 90^0 corresponding to the integral of the curvature function extended over the sequence numbers where the phase is invariant, in good agreement with the experimental localization of the curvature at $n = 135 \pm 5$, and its value derived from experimental data [6].

The biological role of curvature and in general of the sequence dependent DNA superstructure is not yet quite clarified, but it is suggested that, apart from its importance in the mechanism of

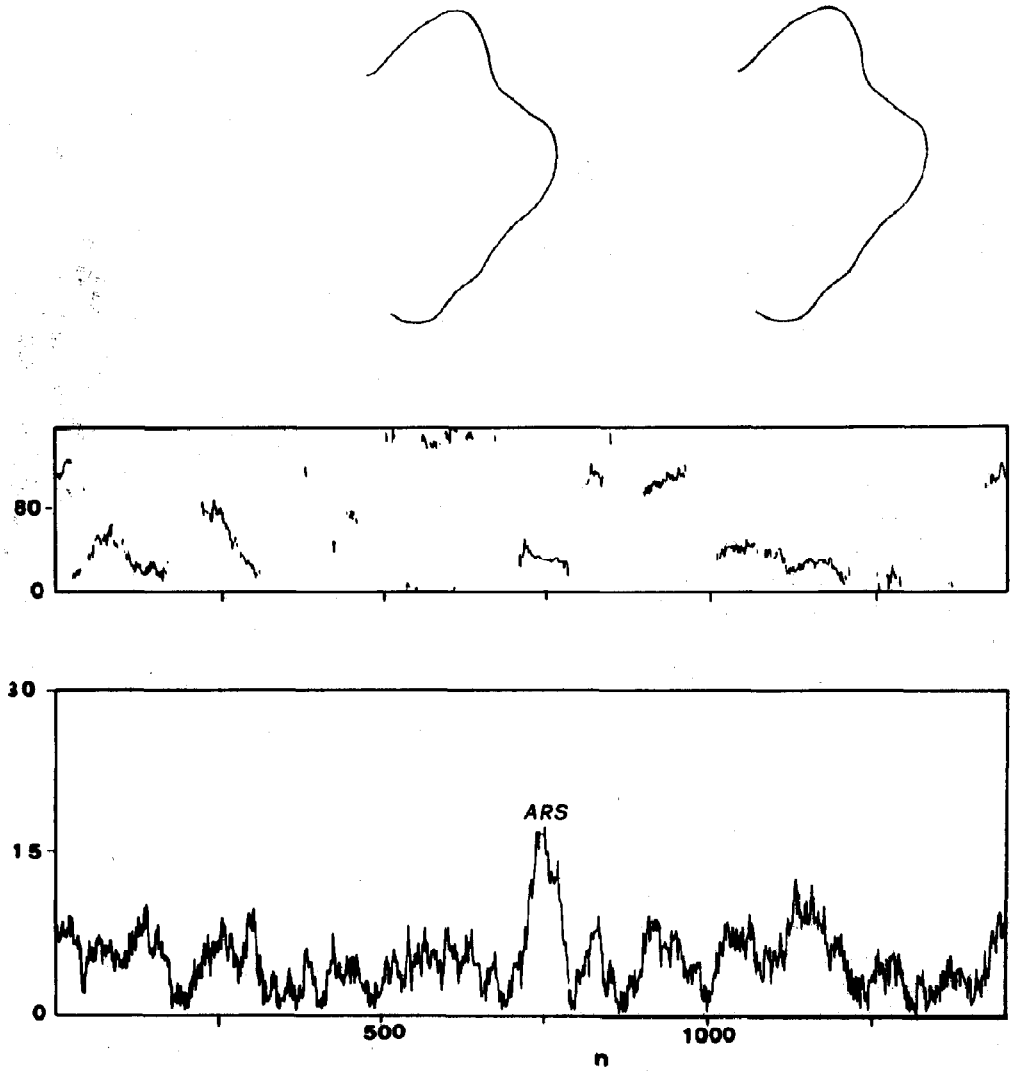


Fig. 5. Curvature and relative phase of a fragment of the TRP1 plasmid. The writhing of the double helix axis is shown in the more extended stereoprojection.

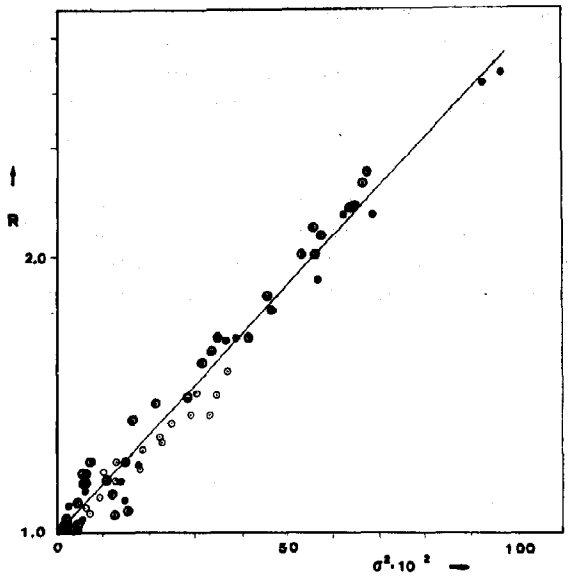


Fig. 6. Linear correlation between the gel electrophoresis retardation and the theoretical values of σ^2 (correlation factor = 0.99). The diagram reports 83 polynucleotides listed in table 1, different either for the sequence or the molecular weight (○, $n = 100$; ⊙, $n = 150$; ●, $n = 200$).

DNA folding in virus capsids and in mitochondria as well in cyclization reactions, it is involved in the topological mechanism of super-twisting DNAs as very recently shown by Laundon and Griffith [33] as well as in phasing of nucleosomes [34]; further there is evidence that the curvature provides a first level of the protein recognition processes in gene regulation [35–38].

Fig. 5 illustrates the curvature diagram of the TRP1-ARS1 plasmid when the ARS (autonomously replicating sequence) region emerges as the highest maximum in agreement with the electrophoretic retardation experiments on TRP1-ARS1 [39]. Retardation was observed, also, at the λ [8] and SV40 [9] DNA replication origins.

Finally, the curvature is involved in the preferential positioning of nucleosomes along DNA, as suggested by the Drew and Travers experiments [34] of differential cleavage by DNAase on 169 bp circularized tract of DNA; they showed that the curvature phase is the same in circular as well as in reconstituted nucleosomes obtained with the same but nicked DNA.

We analyzed the curvature function of a DNA tract about the *Xenopus* 5 S RNA gene, recently investigated by Rhodes [40] using the technique of the differential cleavage, and found a curvature profile characterized by the presence of a quasi mirror plane at the sequence number $n = 100$ in good agreement with the experimental position of nucleosome dyad axis [14]; further, the presence of a curvature minimum about this position agrees with the model of nucleosome recently proposed [41].

6. Structural fluctuations and electrophoretic properties of curved DNAs

Whilst the curvature per turn as well as the associated phase is constant for each periodical sequencial polynucleotide, it is variable in the case of natural DNAs as clearly shown in the curvature diagrams presented. The particular influence on the DNA global writhing and the related physical properties of the phase is very significant: an extensive coherence along the sequence is, in fact, a necessary condition for amplifying the local curvatures at the level of physical detection.

We have found it useful to investigate the dispersion or variance of the local curvature vectors by calculating the moments of the distribution in the complex plane of the deviation vectors d_j. We define the first moment as the average value of $C(n, v) = \sum_{j=n}^{n+v} d_j e^{2\pi i(j-1)/v^0}$ of a given DNA tract between n_2 and n_1 sequence number, namely

$$\langle C \rangle = 1/(n_2 - n_1) \sum_{v=1}^{n_2 - n_2} C(n_1, v)$$

and the relative variance through the second moment

$$\langle C \mid C \rangle = 1/(n_2 - n_1) \sum_{v=1}^{n_2 - n_1} C^*(n_1, v) C(n_1, v)$$

as $\sigma^2 = \langle C \mid C \rangle - \langle C \rangle^2$ which represents an analogy of the square radius of gyration of the deviation vectors chain taking into account the relative

phases. Remembering that the deviation vectors represent the projection on the xy plane (the complex plane) of the unit vectors perpendicular to the base pairs' average plane, σ^2 corresponds to the central dispersion of these projections along the tract of the sequence considered and therefore is a measure of the intensity of snaking of the double helix from the straight DNA.

Therefore σ^2 was expected to show a linear correlation with the electrophoretic retardation factors according to theories of gel electrophoresis [42,43] which predict a dependence of mobility on mean square end-to-end distance of a chain molecule; this quantity is, in the case of a random chain, proportional to the square radius of gyration. It should be noted that the presence of a strong harmonic component in the local structural fluctuation along the sequence in phase with the

Table 1

Experimental electrophoretic retardation factor (R) over the corresponding theoretical dispersion of the structural fluctuations (σ^2) for periodical polynucleotides at different degrees of polymerization (N)

Repeating sequences	$R/\sigma^2 \cdot 10^{-2}$		
	$N=100$	$N=150$	$N=200$
CCGGCAAAAACGGGC	1.01/0.99	1.00/1.02	1.00/1.02
CGGAATTCCG	1.18/11.78	1.46/20.13	
CCTTTAAAGG	1.00/1.50	1.00/2.58	1.00/3.40
GGCCAAAACG	1.29/16.84	1.61/28.81	1.70/35.74
TCCCCGGGGA	1.00/0.27	1.00/0.39	1.00/0.44
CAAAAAAAAA	1.22/17.66	1.70/34.80	
AGGGGCCCCT	1.00/0.28	1.00/0.41	1.00/0.47
CCAAAAACGGGCAAAAAAAA	1.32/21.63	1.70/39.97	1.91/54.56
ACCCCGGGGT	1.00/0.33	1.00/0.49	1.00/0.55
CCGAAAAAGG		1.80/43.94	
TGGGGCCCCA	1.00/0.13	1.00/0.16	1.00/0.17
CCAAAAAAAACGGGCAAAAA	1.39/23.78	1.85/44.21	2.15/60.38
GGGTCGACCC		1.01/0.24	
GGCAAAAACG	1.50/28.45	2.00/50.45	2.15/65.22
TCGTGGGGGC		1.02/2.05	
GGAAATTTCC		2.00/53.94	
GGCAACAACG	1.03/2.10	1.02/3.54	1.04/4.37
GGCAAAAATG		2.07/54.69	
CTTTTAAAAG	1.01/1.11	1.04/1.82	1.09/2.47
CCAAAAAAAA	1.42/28.23	2.10/54.62	
GTTTTAAAAC	1.02/2.14	1.10/3.83	1.14/5.41
GAAAATTTTC	1.42/32.44	2.17/62.89	2.63/91.16
GGCAAGAACG	1.06/6.34	1.13/10.85	1.11/13.46
AAAACGGGTTTTTTGGGCAA		2.18/61.90	
GGCAATAACG	1.08/5.45	1.18/9.61	1.18/12.40
CAAAATTTTG	1.49/34.66	2.26/67.14	2.67/97.30
GGCCAAACCG	1.12/8.04	1.25/13.24	1.24/15.76
GGCAAAAAAC	1.58/35.55	2.30/65.17	
GTGTGGTTAA		1.05/12.28	
CAAAAAACGG	1.17/5.66	1.20/5.90	1.17/5.94
CAATCATTTT		1.40/15.89	
GGCCAAAAACG	1.34/20.58	1.65/31.15	1.69/34.16
AATTTCAAAT		1.25/7.22	
AAACTAGTTT		1.07/14.84	
GGCCAAAAACCG	1.17/5.04	1.20/5.12	1.17/5.05
CGGAATTCCGCGGAATTCCGG	1.25/11.85	1.48/26.65	

period of B-DNA, polarizes the deviation vectors chain in a given direction with the consequence of a higher central dispersion. The value of σ^2 is, in fact, more sensitive to the large component of the second moment which is the rate determining of the electrophoretic mobility. We evaluated σ^2 for the periodical polynucleotides available in literature as well as for tracts of natural DNAs.

Fig. 6 illustrates the surprisingly good linear correlation between the gel electrophoresis retardation factors and the theoretical values of σ^2 (correlation factor $= 0.99$). The diagram reports 83 polynucleotides listed in table 1, different either for the sequence or the molecular weight. It also predicts correctly the dependence of the retardation factor from the molecular weight which in the case of periodical polynucleotides shows a flat maximum, when a spire of superhelix is completed, followed by a slight decrease.

Note that a very small (2%) but significant improvement of the correlation factor was obtained using an average azimuthal angle ϕ evaluated as a function of the base composition ($\phi^0 = 34.6^0 + x_{AT}$; x_{AT} being the molar fraction of A + T) of the sequence investigated on account of the observed increasing of the twist angle in AT-rich DNAs.

7. Theoretical prediction of the permutation gel electrophoresis from the sequence

The strikingly good linear correlation obtained between the gel electrophoretic retardations and the theoretical variance of the structural deviations from the standard B-DNA structure, valid for any sequence in a large range of molecular weight, prompted us to extend such a model to fragments of natural DNAs. In fact, the model allowed the reproduction of the gel permutation experiments of different fragments of DNAs experimentally investigated by different authors.

Fig. 7 illustrates the cases of the fragments of kinetoplast DNA from *Leishmania tarentolae* which represent the first examples of the technique of permutation gel electrophoresis for localizing the DNA curvature, as introduced by Wu and Crothers [6]. The diagrams report the calculated variance σ^2 of 1–423 and 1–241 tracts of kDNA as obtained by a cyclic permutation of the sequence; they appear practically identical, also for the fine features of the gel permutation diagrams (figs. 2 and 3 of the already cited authors [6]). Thus, the theoretical diagrams of some fragments of SV40, very recently investigated by Milton and Gesteland [44] by gel electrophoresis and by Hsieh and Griffith by electron microscopy [45], very satisfactorily fit the experimental gel mobility profiles.

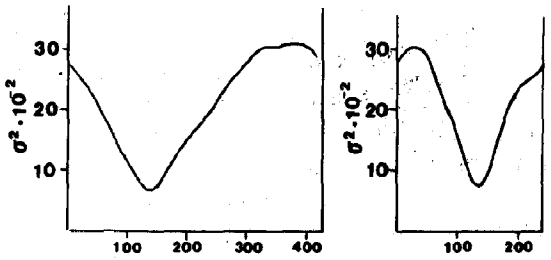


Fig. 7. Diagrams of the calculated variance σ^2 of 1–423 (left) and 1–241 (right) tracts of kDNA.

Fig. 8 illustrates the case of the plasmid pT 181 investigated by Koepsel and Khan [46] using the same technique of relative gel electrophoresis

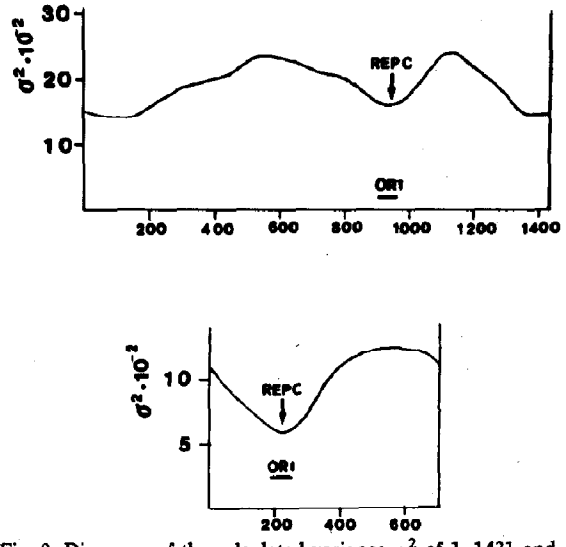


Fig. 8. Diagrams of the calculated variance σ^2 of 1–1431 and 721–1431 pT181 fragments.

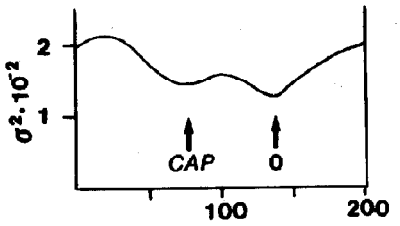


Fig. 9. σ^2 permutation diagram of DNA binding domain of *E. coli* CAP protein.

migration of permutated fragments. They gave evidence of DNA bending at the replication origin. In perfect agreement with their experimental data, the σ^2 diagrams along the permutating sequence of the 1–1431 and the 721–1431 fragments of pT 181, in spite of the large molecular weight, clearly show a minimum at the same sequence number corresponding to the origin of replication where the replication initiator protein, RepC, binds. This result is particularly significant because, although the average base composition of pT181 is about 30% G + C, the origin region contains about 50% G + C base pairs, in contrast with the approximate notion that curved DNA should be associated only with A − T rich DNA sequences. The ability of our model to theoretically localize the DNA curvature, as a good alternative to gel electrophoresis experiments, is better appreciated since the sensitivities of the two techniques are compared.

Fig. 9 shows, in fact, the σ^2 permutation diagram of DNA binding domain of *E. coli* CAP protein recently investigated by electrophoretic analysis [47]. As demonstrated by Wu and Crothers [6], the CAP protein binds to the *lac* promoter region at the position −62 from the start site of transcription as shown by the presence of a retardation minimum in the permutation gel electrophoresis of the CAP-DNA complex; the naked DNA fragment, however, did not show significant deviation from a uniform mobility.

We evaluated the σ^2 diagram which, in spite of its low profile, clearly showed the position of the CAP protein together with another minimum localized at the start site of transcription where RNA polymerase binds. Finally, as further examples fig. 10 shows the presence of a minimum at the dyad axis of the nucleosome on *Xenopus* 5 S gene as well as at the locus of the terminal transcription of SV40.

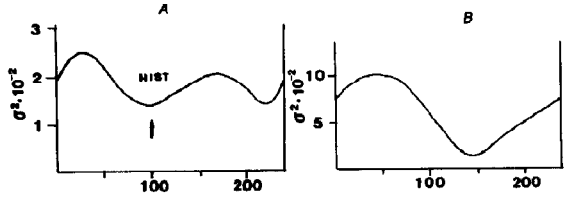


Fig. 10. σ^2 permutation diagram of (A) *Xenopus* 5 S gene fragment; (B) SV40 2461–2699 terminal transcription region.

8. Concluding remarks

The theoretical model presented in this paper certainly requires further investigations and refinements; neverthless, it appears to contain the basic aspect of the problem for translating the sequence deterministic fluctuations into superstructural elements of DNA.

This important ability of DNA to integrate the differential stereochemical parameters of different nucleotides, giving rise to sequence dependent superstructures, is not quite evaluated, but it is clear that it could provide an insight into the molecular mechanisms of the management of the genetic information encoded in DNA.

The formal simplicity of the model presented allows its extensive use for an easy screening of sequence dependent curvature in DNAs and the results obtained seem to assure a good reliability of the theoretical localization of DNA regions which are possible candidates for protein recognition.

Acknowledgments

We thank Miss Grazia Gallo for cooperation in the computations. The work was supported by the Progetto Strategico Biotecnologie CNR.

Appendix

A1. DNA curvature vector from the transformation matrices

The representation of the DNA superstructure in terms of local deviations from the standard B-DNA structure is conveniently given in terms of transformation matrices which relate the orientation of the average base pair plane at a generic jth position of the sequence to that fixed at the origin of the coordinate system.

Let $A_j = \Phi_j R_j T_j$ be the product of the pure rotation matrices about the z, y and x axes by the azimuthal angle ϕ_j, the roll angle r_j and the tilt angle t_j, respectively; it transforms the coordinate system fixed at the the jth base pair, as shown in fig. 1, into that, equivalently fixed, at the next one. Thus, $A(n, v) = \prod_{j=n}^{n+v} A_j$, transforms the nth into the $(n+v)$th coordinate system and the corresponding transpose matrix $A^+(n, v)$ transforms back the $(n+v)$th base pair into the nth one. The smallness of the roll and tilt values and of the deviations of the ϕ_j from their average ϕ^0 value allows the reduction of the matrix product to the following simple form in the first order of the pertinent Taylor series:

$$A^+(n, v) = \begin{vmatrix} \Delta^+(v) & C(n, v) \\ -[\Delta(v)C(n, v)]^+ & 1 \end{vmatrix}$$

where:

$$\Delta(v) = \begin{vmatrix} \cos\delta(v) & \sin\delta(v) \\ -\sin\delta(v) & \cos\delta(v) \end{vmatrix}$$

$\delta(v)$ representing the deviation from 2π of the sum of the azimuthal angles between the $(n+v)$th and nth nucleotides; practically $\delta(v) = v\phi^0 - (\mathrm{mod}2\pi)$; and $C(n, v)$ is the curvature vector (in rad) assigned at the position n of the sequence as a result of the local deviations d_j of the base pairs between the $(n+v)$ and n sequence numbers:

$$C(n, v) = \sum_{j=n}^{n+v} (\Phi^{j-1})^+ d_j$$

where

$$d_j = \begin{vmatrix} r_j \\ -t_j \end{vmatrix}$$

is the deviation vector; the curvature vector represents the angular deviations of the double helix axis between the sequence number n and $n+v$. The curvature vector and the phase $\delta(v)$ contain the information on the superstructure of DNA. It is convenient to choose v about a simple multiple of v^0, the periodicity of the B-DNA structure that different experimental data indicate to be about 10.4 corresponding to the value of $\phi^0 = 34.6^0$. We define the curvature vector per turn $C(n) = (v^0/v)C(n, v)$ and the corresponding phase $\delta = (v^0/v)\delta(v)$. $C(n)$ represents the average deviations of the helical axis per turn of B-DNA. We have earlier shown [14] that in the case of periodical polynucleotides it is easy to define from the elements of the transformation matrix $A^+(n, v)$ the superhelical axis

$$H = \begin{vmatrix} a_{32} - a_{23} \\ a_{13} - a_{31} \\ a_{21} - a_{12} \end{vmatrix} / 2\sin\omega = \begin{vmatrix} -C_y \\ C_x \\ \sin\delta \end{vmatrix} / \sin\omega$$

where $\sin\omega = [|C|^2 + \sin^2\delta]^{1/2}$, in terms of the components of the curvature vector; it is independent from n when v is chosen as a multiple of the sequence periodicity. Thus, it is easy to obtain the superhelical pitch $P = 34\sin\delta/\sin\omega$ (Å) and the superhelix curvature radius $R = 34/|C|$ (Å). The screw sense of the superhelix will be that according to the sign of δ: e.g. it is negative for the 10-fold periodical polynucleotides and positive in the case of the 11-fold polynucleotides. Finally, the grooves as evaluated at the level of phosphate groups are reduced for the fraction 0.3 $C(n)$ in the curvature vector direction. The curvature vector can be conveniently calculated in the complex plane by associating the x and y axes to the real and imaginary axes respectively; thus the vectorial sum transforms in the algebraic sum:

$$C(n, v) = v^0/v \sum_{j=n}^{n+v} d_j \exp[2\pi i(j-1)/v^0]$$

where $d_j = r_j - it_j$; *it* represents the amplitude corresponding to the B-DNA structure periodicity, v^0, of the Fourier transform of the local structural fluctuation along the sequence, d_j. If v is chosen about a multiple of v^0, $C(n, v)$ is about

zero unless the structural fluctuations contain a harmonic component of period v^0. It is particularly important that the curvature is invariant if a constant value is added to all d_j, namely, if the tilt and roll values associated with the different dinucleotides are changed by a constant.

References

1 E.N. Trifonov, Nucleic Acids Res. 8 (1980) 4041.
2 E.N. Trifonov, Cold spring Harbor Symp. Quant. Biol. XLVII (1983) 271.
3 H.C.M. Nelson, J.T. Finch, B.F. Luisi and A. Klug, Nature 330 (1987) 221.
4 M. Coll, C.A. Frederick, A.H. Wang and A. Rich, Proc. Natl. Acad. Sci. USA 84 (1987) 8385.
5 J.C. Marini, S.D. Levene, D.M. Crothers and P.T. Englund, Proc. Natl. Acad. Sci USA 79 (1982) 7664.
6 H.M. Wu and D.M. Crothers, Nature 308 (1984) 509.
7 S. Diekman and J.C. Wang, J. Mol. Biol. 186 (1985) 1.
8 K. Zahn and F.R. Blattner, Science 236 (1987) 416.
9 K. Ryder, S. Silver, A. Lucia, E. Fanning and P. Tegtmeyer, Cell 44 (1986) 719.
10 G.M. Clore and A.M. Gronenborn, FEBS Lett. 179 (1985) 187.
11 M.H. Sarma, G. Gupta and H.R. Sarma, Biochemistry 27 (1988) 3423.
12 V.B. Zurkin, Y.P. Lysov and V.J. Ivanov, Nucleic Acids Res. 6 (1979) 1081.
13 A.R. Srinivasan, R. Torres, W. Clark and W.K. Olson, J. Biomed. Struct. Dyn. 5 (1987) 459.
14 P. De Santis, S. Morosetti, A. Palleschi and M. Savino, in: Structure and dynamics of nucleic acids, proteins and membranes, vol. 31, eds. E. Clementi and S. Chin (Plenum Publishing, New York, 1986).
15 P. De Santis, S. Morosetti, A. Palleschi, M. Savino and A. Scipioni, in: Biological and artificial intelligence systems, eds. E. Clementi and S. Chin (ESCON Science Publishers B.V., 1988).
16 H.S. Koo, H.M. Wu and D.M Crothers, Nature 320 (1986) 501.
17 P.S. Hagerman, Nature 321 (1986) 449.
18 S. Diekmann, EMBO J. 6 (1987) 4213.
19 H.S. Koo and D.M. Crothers, Proc. Natl. Acad. Sci USA 85 (1988) 1763.
20 S. Arnott, R. Chandrasekaran, I.H. Hall and L.C. Puigjaner, Nucleic Acids Res. 11 (1983) 4141.
21 J. Griffith, M. Bleyman, C.A. Rauch, P.A. Kitchin and P.T. England, Cell 46 (1986) 717.
22 P. De Santis, E. Giglio, A.M. Liquori and A. Ripamonti, J. Polymer Sci. A 1 (1963) 1383.
23 P. De Santis, E. Giglio, A.M. Liquori and A. Ripamonti, Nature 206 (1965) 456.
24 F. Calascibetta, M. Dentini, P. De Santis and S. Morosetti, Biopolymers 14 (1975) 1667.
25 P. De Santis, S. Morosetti, A. Palleschi and M. Savino, Biopolymers 20 (1981) 1707.
26 P. De Santis, S. Morosetti, A. Palleschi and M. Savino, Biopolymers 20 (1981) 1727.
27 P. De Santis, R. Rizzo and G. Ughetto, Nature 237 (1972) 94.
28 P. De Santis and A.M. Liquori, Biopolymers 10 (1971) 699.
29 C.R. Calladine, J. Mol. Biol. 161 (1982) 343.
30 C. Yoon, G.G. Privé, D.S. Goodsell and R. Dickerson, Proc. Natl. Acad. Sci. U.S.A. 85 (1988) 6332.
31 P. De Santis, S. Morosetti, A. Palleschi, M. Savino and A. Scipioni, in: Proceeding of workshop on Molecular Structure Dynamics and Recognition, Siena, Italy, 1988 (in press).
32 D.S. Ray, J.C. Hines, H. Sugisaki and C. Sheline, Nucleic Acids Res. 14 (1986) 7953.
33 C.H. Laundon and J. Griffith, Cell 52 (1988) 545.
34 H.R. Drew and A.A. Travers, J. Mol. Biol. 186 (1985) 773.
35 L. Bossi and D.M. Smith, Cell 39 (1984) 643.
36 R.R. Plaskon and R.W. Wartell, Nucleic Acids Res. 15 (1987) 785.
37 J.N. Anderson, Nucleic Acids Res. 14 (1986) 8513.
38 K. Zahn and F.R. Blattner, Nature 317 (1985) 451.
39 M. Snyder, A.R. Buchman and R.W. Davis, Nature 324 (1986) 87.
40 D. Rhodes, EMBO J. 4 (1985) 3473.
41 S.C. Stachwell, H.R. Drew and A.A. Travers, J. Mol. Biol. 191 (1986) 659.
42 L.S. Lerman and H.L. Frisch, Biopolymers 21 (1982) 995.
43 O.J. Lumpkin and B.H. Zimm, Biopolymers 21 (1982) 2315.
44 D.L. Milton and R.F., Nucleic Acids Res 16 (1988) 3931.
45 C.H. Hsieh and J.O. Griffith, Cell 52 (1988).
46 R.R. Koepsel and S.A. Khan, Science 233 (1985) 1315.
47 H.L. Johnson, M.R. Gartenberg and D.M. Crothers, Cell 47 (1986) 995.